LANGUAGE ACQUISITION AS LEARNING

MIKIKO NISHIKIMI, HIDEYUKI NAKASHIMA AND HITOSHI MATSUBARA
Electrotechnical Laboratory
Tsukuba, Japan
nisikimi@etl.go.jp, nakashim@etl.go.jp, matsubar@etl.go.jp

Abstract

Chomsky's proposition that language is handled by a language-specific faculty needs more justification. In language acquisition in particular, it is still in question whether the faculty is necessary or not. We succeeded in explaining one constraint on language acquisition in terms of a general learning mechanism. This paper describes a machine learning system Rhea applied to the domain of language acquisition and shows that Rhea can learn the tendency which children confronting new words seem to have.

1 Introduction

Chomsky proposed that language is handled by a language specific faculty, but this proposition has not been verified, especially in the area of language acquisition. Although Berwick[1] showed the existence of a special mechanism sufficient for the learning of syntax, there is still a question of whether or not the mechanism is necessary. Furthermore, his model does not explain acquisition of semantics or concepts. These were simply presupposed.

We started from a general learning mechanism and succeeded in explaining a constraint on language acquisition.

Children learning their first language face and solve a big problem of induction. They find out how words are used and related to other words from limited information at a surprisingly rapid rate. In the field of developmental psychology, many kinds of constraints have been proposed to account for this phenomenon. Most of these constraints come from the view that assumes a specific framework for language acquisition, but there is another view: language as an extension of other intellectual faculties, and its acquisition as one result of the universal learning process that leads to our acquisition of intellect.

We want to explain the children's ability in terms of the latter view. Thus, we make a machine learning system, Rhea, which accepts n-tuple inputs consisting of instances from n-domains (one from each domain) and creates the rules that delimit the possible combinations. This framework is very general, and yet if we choose outer-worlds and linguistic descriptions for them as two input domains, it can be seen as a language acquisition system without language-specific constraints.

In this paper, we describe the machine-learning system Rhea and its application to the domain of language acquisition. We show that without a priori information about how outer-worlds are organized, Rhea can learn the "setting for new words", which children confronting new words seem to possess.

The point is how the model acquires and formalizes the "meaning" of an expression. To achieve this autonomously, Rhea has its own representation language for outer-worlds. If one linguistic expression is repeatedly given along with different outer-worlds, it builds up one common representation for all the outer-worlds. This internal representation that has a one-to-one correspondence to a linguistic expression is regarded as the "meaning" of the expression in our model.

2 Constraints

In order to elucidate the children's rapid acquisition of vocabulary, constraints on the possible hypotheses about the meanings of linguistic expressions have been postulated. Clark[2] proposes *the principle of contrast* whereby every two forms contrast in meaning, and Markman[3] suggests a stronger *assumption of taxonomic organization.*

The assumption of taxonomic organization confines children to assuming that a word given with an unknown object refers to a taxonomic class of the object. As ostensive definition is the only way to acquire early vocabulary, the assumption reduces the possible search space of meaning. With this assumption, if you see someone point to an unfamiliar object and say a word, you can presume that the word is either the label of the object or the label of one of the categories it belongs to and can forget about the possibility of the word's referring to one of its attributes or

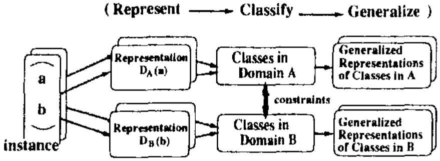

Figure 1: Information flow

its relation to other objects.

Children seem to consider the assumption of taxonomic organization. Markman's experiment shows that even though they are liable to consider thematic relations in domains other than language acquisition, children hearing a new word attend to taxonomic relations. This tendency is called the "setting for new words".

It is not clear, however, if such constraints are innate or not, or more essentially if they can be derived from restrictions that any intelligent system should observe. One way to clarify this point is to examine whether the model that does not contain the constraint can acquire it during the learning process.

3 An overview of Rhea

3.1 Rhea as a machine learning system

Fig.1 illustrates Rhea's learning process in two different domains, A and B. The system's task is to find general rules that predict which instance from Domain A can appear with a certain instance from B, and vice versa.

Rhea accepts as input a pair of instances $i = \langle a, b\rangle$. One instance is from Domain A and the other from Domain B. One pair is given at a time. Rhea is equipped with an internal representation language for each domain, D_A and D_B, and has predefined methods to extend the representation languages in case of need. Similarities, generalization operations and specialization operations are defined upon each language. Rhea represents an input pair using these languages and their extensions, and makes an internal representation $D(i) = \langle D_A(a), D_B(b)\rangle$, which is a pair of a representation of Domain A instance and that of a Domain B instance. More than one possible internal representation may exist for one input, but the one found first is stored.

When representations are accumulated, Rhea is able to find out rules. It first sorts internal representations into classes based on similarities. Classes may or may not overlap. Then Rhea generalizes representations of each class. This process of classification and generalization is done on demand.

When a partial input a (an instance from Domain A) is given and its counterpart b (from Domain B) is to be predicted, the model first classifies the partial input into a class N using the information about a, makes the generalization of Domain B part of all the other representations in class N and expects one of its specializations to be b's representation $D_B(b)$.

The model forms classes so that representations in each class share some characteristics. Two internal representations, $\langle D_A(a_1), D_B(b_1)\rangle$ and $\langle D_A(a_2), D_B(b_2)\rangle$, belong to the same class if $D_A(a_1)$ and $D_A(a_2)$ are similar in the criterion defined in the representation language D_A, and $D_B(b_1)$ and $D_B(b_2)$ are also similar in the criterion defined in D_B. In the extreme case, if $D_A(a_1)$ equals $D_A(a_2)$, then $D_B(b_1)$ must equal $D_B(b_2)$ and vice versa, which means that when two instances from one domain are represented as the same, instances from the other domain that appear with them must also have the same internal representation.

3.2 Rhea as a language acquisition model

Rhea, when applied to the domain of outer-worlds S and the domain of linguistic expressions L that describe the outer-worlds, can be regarded as a language acquisition model.

In these domains, Rhea learns the followings:

1. Extensions of the representation language of linguistic expressions D_L
2. Internal representations of linguistic expressions $D_L(l_1), \ldots, D_L(l_n)$
3. Extensions of the representation language of outer-worlds D_S
4. Internal representations of outer-worlds $D_S(s_1), \ldots, D_S(s_n)$
5. Classification of inputs

which respectively can be seen as

1. Syntactic rules
2. Structures of linguistic expressions
3. Concepts that delineate meanings
4. Meanings of linguistic expressions derived from outer-worlds

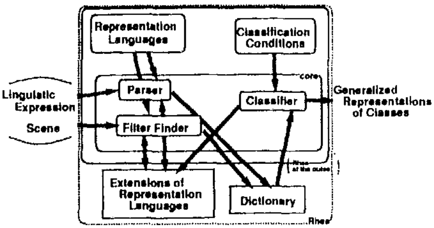

Figure 2: Rhea as a language acquisition model

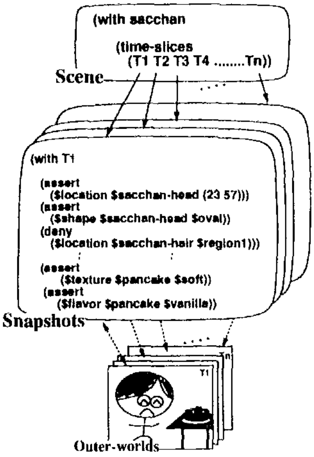

Figure 3: Scene: a part of the input

5. Categories of linguistic expressions.

Fig.2 shows the configuration of the language acquisition model, Rhea. It receives a pair of one *scene* and a linguistic expression that describes the scene. An expression is a sequence of words and contains no structural information. A scene is the equivalent of sensory input from outer-worlds. Fig.3 shows an example of a scene. A scene is a sequence of *snapshots* which are lists of assertions that become true or false at the time when the snapshots have been taken. Each assertion expresses a relation between two terms. The terms may be objects, attributes or values, which cannot be distinguished by Rhea.

The *parser* makes the internal representations of linguistic expressions, and the *filter finder* makes those of scenes. The *classifier* divides representations into classes and makes rules. Since two inputs represented as the same in one domain must have the same representation in the other domain, there may be no synonyms or polysemants, which means that the model has "the principle of contrast" implanted from the beginning.

4 Internal representations of inputs

The internal representation of an input is a pair of internal representations of the input's constituents, which is a pair of one *structure* and one *filter*.

4.1 Internal representation of linguistic expressions

The internal representation of a linguistic expression is the syntactic structure of the expression. For example, a linguistic expression `''Kitty ate pancakes''` is internally represented as a *structure*

```
S: (Sentence (Class1 'Kitty')
             (Class2 (Class3 'ate')
                     (Class4 'pancakes')))
```

The first element in the list specifies the name of the class the structure belongs to and the rest are its constituents. Each constituent in turn has its class name and constituents.

The representation language D_L at the beginning contains suppositions that one input expression forms one structure and can be described with a phrase structure grammar. The model accepts a new input expression provided that it can be described by adding at most one new rule. When known rules cannot parse an expression, Rhea parses it from the bottom to up and from the top to down simultaneously and makes partial structures. If they can be combined into one structure by adding one rule, Rhea adds the rule to the memory as an extension of D_L. If one rule cannot connect all of them, the model backtracks to find another parsing or abandons the input.

Rhea sets the class of an unknown word considering the scene given with the word. When some rule predicts the class of the word and the scene presented with the word can be given an internal representation similar to those of other words in the class, the word is added to the predicted class. If not, a new category is assigned to the word.

4.2 Internal representation of scenes

An internal representation of a scene provides the semantics of the linguistic expression that comes

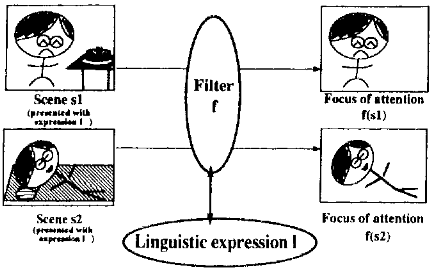

Figure 4: Relationship among Filter, Scenes and Focus of Attentions

with the scene. Linguistic expressions change or control the listeners' interpretations of the outer-world, and make speakers and listeners share one *focus of attention* (hereinafter, *FOA*). In order to model this process, a scene is internally represented as a procedure that converts the scene into an FOA. We call this procedure a *filter*. As stated before, a scene is a sequence of lists of assertions, and so is an FOA. FOAs must contain at least one non-variable assertion because there must exist non-variable FOAs to be shared among speakers and listeners. If a filter applied to scene s yields a non-variable sequence of lists of assertions, the filter is *valid* for s.

Any valid filter for scene s can be a representation of the scene. For example, a scene that contains someone eating pancakes may be internally represented in several ways. A procedure that focuses the listeners' attention on pancakes and yields pancakes as an FOA is valid for the scene, and one that stresses the eating action can also be an internal representation of the scene. However, scenes which appeared with the same expression must have the same filter because there may be no polysemants.

Fig. 4 shows the relationship among filters, scenes and FOAs. Since the FOAs derived by filter f from scene $s1$ and scene $s2$ both contain some objects, the filter is valid for both scenes. Thus two scenes that appear with linguistic expression l are represented by the filter.

4.2.1 Representation language of filters

Filters are mappings from scenes to FOAs. Rhea has 32 parameterized simple mappings as its representation language D_S at the start. It combines mappings and searches a given scene for values to instantiate parameters. We call these parameterized mappings *filter-primitives* and instantiated mappings *filter-elements*. For instance, among the possible combinations of filter-primitives is the one

```
(snap-remove not-include *variable*)
```

which removes assertions that do not contain a certain term from a snapshot. When a scene is given, the model selects one of the terms in the scene, namely `$location`, to substitute for `*variable*` and makes a filter-element

```
(snap-remove not-include $location)
```

which extracts assertions that contain the term `$location` from a snapshot in the scene. A filter is a sequence of one or more filter-elements. Filter-elements in the sequence are applied to a scene one by one and the result becomes the FOA.

4.2.2 Acquisition of filters

Rhea shapes filters through trial and error. Whenever a new scene is given with an expression, the filter that seems to correspond to the expression is tested for its validity for the new scene, and Rhea then elaborates or corrects the filter depending on the result.

When the new input $\langle l, s\rangle$ is given, the model creates $D_L(l)$, which is the representation of l by the language D_L, and searches through the memory for a representation that has the form $\langle D_L(l), f\rangle$, where f is an internal representation of an instance from Domain S.

If there is no representation of the form $\langle D_L(l), f\rangle$, l is regarded as a new expression and Rhea builds a candidate for filter f. The candidate consists of one filter-element made by selecting one filter-primitive randomly and substituting terms in the given scene s for parameters of the filter-primitive. If the candidate is valid for scene s, it is used as an internal representation of the scene. If it is not, another candidate is created and tested. As there must be no synonyms, a filter must be different from those of other expressions.

If Rhea already knows the linguistic expression l, that is, if the representation of the form $\langle D_L(l), f\rangle$ is in the memory of Rhea, it checks the validity of filter f for scene s. Rhea elaborates valid filters and corrects invalid ones.

Elaboration is to make filters more specific by adding conditions. Rhea may either insert one randomly selected filter-element into the existing filter or replace one filter-element by a more specific one. For each input, the model can add only one condition, so learning proceeds gradually. The new filter must be different from the filters of other expressions and must extract an FOA which is different from the one derived by the old filter. If Rhea cannot elaborate the filter to make up a new one, it keeps the old one.

Correction of a filter is done by deleting conditions. Rhea keeps a *revision counter* R for every internal representation. It is the number of successive scenes from which the filter cannot extract an FOA and Rhea cannot correct it. To correct a filter, Rhea may remove j filter-elements, replace parameters of k filter-elements with other values extracted from scene s or replace l filter-elements with more general ones. The number of changes $j+k+l$, however, must not exceed the value of the revision counter. When the correction succeeds, Rhea sets the revision counter to zero. If the filter cannot be made valid for scene s within the allowed number of changes, Rhea keeps it and increments the revision counter by one.

5 Classification and generalization of input

Rhea divides internal representations into classes. A class contains representations that have both similar structures and similar filters. As classes may overlap, an internal representation can be a member of two or more classes.

5.1 Similarity of structures

Two structures are similar if they are in interchangeable positions within bigger structures. For example, having two structures:

```
S1: (Sentence (Category1 'yellow')
              (Category2 'pancake'))
S2: (Sentence (Category1 'red')
              (Category3 'raspberries'))
```

may trigger the making of a class that contains two representations whose structures are `(Category2 'pancake')` and `(Category3 'raspberries')` respectively. These structures are *similar* because they both have one `Category1` as their sister class and form members of the `Sentence` class.

5.2 Similarity of filters

Filters are lists of filter-elements. Two filters are similar when they can be *generalized* into the same non-null and non-variable list. Rhea has the following generalization (= dropping down conditions) operations.

1. deletion or transformation into a variable of a filter-element at a specified position in the list
2. deletion or transformation into a variable of filter-elements between those that match certain patterns
3. transformation into a variable of a part of a filter-element at a specified position in the list

If a sequence of operations is applied to a set of filters and yields a common and non-trivial result, the internal representations that have those filters can form one class.

For example, an internal representation with a filter `((F x y) (G v))` and another representation whose filter is `((F x z))` may belong to the same class because the non-trivial generalization of the two filters `((F x *variable))` exists.

5.3 How classes can be used

As described in subsection 3.1, a class constrains its members to a certain form of representation. There are two ways for the model to use this restriction.

One way is based on the class-instance relations among representations. We can demarcate the search space for the meaning of the expression if its class is known.

Rhea, in need of finding the filter paired with a structure, first determines the class of the structure, generalizes all the filters of members of the class and expects that the filter in question is one of the specializations of the generalized filters. Specialization is done by substituting values for variables in the generalized filter or adding one or more filter-elements to the filter.

The other way utilizes meta-relationships of relationships among representations. The structures define whole-part relationships among themselves. Representations of a class are expected to share some characteristics of these relationships. We can guess the meaning of a sentence that was never heard before. This happens when we know all the constituent words and how their meanings contribute to the meaning of the whole sentence.

When a new linguistic expression is given and represented in a structure, Rhea can accelerate the search for the filter paired with it if the filters of its constituents are known. It first identifies the structure's class, and then makes one rule for each member of the class that explains how the filter of the member is broken down into the

Table 1: Possible forms of input sentences

```
<S> ::= [<N>][<N>][<V>]
<N> ::= <a><N> | <n> | <p>
<V> ::= <v> | <a>
<n> ::= "asi" | "atama" | "ahiru" | "okasi"
      | "cup" | "kuti" | "glass" | "coffee"
      | "sara" | "spoon" | "tabemono"
      | "tukue" | "te" | "ikimono" | "neko"
      | "pancake" | "milk" | "me"
<p>::=  "Kitty" | "Sacchan" | "Huey"
      | "Dewey" | "Louie" |
<v> ::= "aru" | "ugoku" |"sawaru"
      | "taberu" | "nai"
<a> ::= "kiiroi" | "amai" | "kuroi"
      | "marui"
```

filters of its constituents. It then generalizes all these rules and expects that a specialization of the generalized rule applies to the structure in question. Therefore, Rhea puts the filter of its constituents into the general rule and composes a candidate for the filter of the whole structure. The model can limit the search space for the filter to specializations of the candidate.

6 Experiment: one-word sentence

We test the model to see whether it can acquire the "setting" for new words given as one-word sentences.

An input scene is selected from 48 possibilities that we have prepared. The lexicon has 32 words, but not every word can describe a given scene, thus for each scene we made a list of words that can be used to describe it. Linguistic expressions are randomly composed using the words in the list and the grammar shown in Table 1, [1] and are restricted to no more than a length of three words. These **<n>**, **<p>**, **<v>** and **<a>** roughly correspond to nouns, proper-nouns, verbs and adjectives.

After 432 pairs were input, Rhea divided 32 words into three, unconnected classes: Class1, Class2 and Class3. In the internal representations of two- or three-word sentences, they were further divided into subclasses, but here for simplicity, we concentrate on the classes made to express one-word sentences. Class1 contained one **<v>** word "aru" (to exist), Class2 contained another **<v>** word "nai" (not to exist) and all other 30 words were classified into the last class, Class3.

[1] English translations of terminal symbols in Table 1 are:

```
<n> ::= "leg" | "head" | "duck" | "sweets"
      | "cup" | "mouth" | "glass" | "coffee"
      | "plate" | "spoon" | "food"
      | "table" | "arm" | "living thing" | "cat"
      | "pancake" | "milk" | "eye"
<v> ::= "to exist" | "to move" |"to touch"
      | "to eat" | "not to exist"
<a> ::= "yellow" | "sweet" | "black"
      | "round"
```

```
((subseq 0 0)
 (snap-count all)
 (snap-sort all maxcount)
 (map snap-remove not-include *variable*))
```

Figure 5: The general filter of one-word sentences

Rhea learned that the word in Class1 is associated with a filter that extracts assertions that become true at the time of utterance, and the filters of the word in Class2 extracts only assertions that become false at utterance.

Fig.5 shows the generalized filter of Class3. It makes parameterized modifications to scenes. The first filter-element `(subseq 0 0)` extracts changes at the time of utterance, `(snap-count all)` counts how many times each term appears in the snapshot and `(snap-sort all maxcount)` changes order of assertions in the snapshot so that assertions that contains the term that appears more frequently come earlier. The last filter-element `(map snap-remove not-include *variable*)` has a variable and Rhea has to select a term from the snapshot to substitute for it. The substituted filter-element extracts assertions that contain the term. As the assertions in the snapshot are thus sorted, the term that appears most frequently is selected first, and the filter that focuses on the term is tested for its validity first.

As for the relationship between a one-word sentence and its only constituent word, Rhea conjectured that the filter of the sentence is the same as that of the word.

In short, Rhea acquired the general filter for a group of one-word sentences and it extracts such assertions that describe a term that appears most frequently in the snapshot at the time of utterance. As Rhea backtracks, assertions with the next most frequent term are extracted.

Scenes have more labels for an object than labels for its attributes because each assertion expresses a relation between two terms and an object label appears in all the assertions about its attributes. Therefore when the model is given a one-word sentence whose constituent word does not belong to classes of words of existence/nonexistence, it first assumes the sentence to refer to the label for an object in the scene. If the label is already known, the model then backtracks to refer to the label for its most salient attribute or a label for another object. This is what children with the "setting for new words" would do facing a new one-word sentence.

7 Discussion

7.1 Semantic concepts and input

Other acquisition models that cover semantic acquisition are the system of Takagi et. al. [4], which accepts a sentence and visual input, Hill's language acquisition model[5] and Selfridge's CHILD[6]. However these models assume semantic concepts from the start, and their task is to associate linguistic entities with them. These systems, which receive a semantic concept to be associated with a linguistic expression as direct input, cannot misunderstand the meaning of a linguistic expression and cannot shed light on the difficulty of learning the meaning of a certain expression.

We do not assume semantic concepts in representing scenes given to Rhea. We formalize concepts as functions from the direct input to FOAs. They must be formed and tested in accordance with expressions and other concepts. We equipped the model with filter-primitives, which are means of establishing the concepts. We have designed filter-primitives to become equivalents of human abilities of recognition. Filter-primitives are given from the beginning because human beings have the ability to focus their attention on objects, attributes or changes when they begin language acquisition. Rhea can select a parameter from scenes and make concrete filter-elements just like any child coming to distinguish important features in its world. Therefore, our formalization of concepts and its acquisition process is a more realistic one.

7.2 Acquisition of a constraint

The principle of contrast is derived from the general constraint on how a class should be formed to make useful predictions, and as shown in section 6, Rhea has no language-specific constraints but yet can acquire the "setting for new words", because its filter-primitives and classification criteria can reproduce the tendency that was contained in the input pairs.

In our experiment, the one-word sentences given to Rhea were often taxonomic terms or attributes of any objects in the scene and Rhea learned that the best conjecture is that the one-word sentence presented with unknown objects refers to a taxonomic term of the most frequently described object. If we give a label for the biggest object in the scene whenever Rhea meets a scene with multiple objects that are not yet labeled, Rhea will make a filter of a category that sorts objects by size and extracts the first one. Our claim is that children can also acquire the "setting for the new words" from a few inputs of one-word sentences, and that it need not to be set a priori.

8 Conclusion

This paper has described Rhea, the model of language acquisition, which uses very general acquisition procedure. We assume neither semantic concepts nor syntactic rules a priori. Instead, we have equipped the model with the general framework to create the rules that delimit the possible combinations of the input. We applied the model to the domains of outer-worlds and linguistic descriptions of them. The system successfully made concepts that are consistent with given inputs. The experiment showed that it reproduced the "setting for the new words," a human tendency in language acquisition, without language-specific constraints or information about how outer-worlds are organized.

References

[1] R. C. Berwick & A. Weinberg (1983): "*The grammatical basis of linguistic performance*", MIT Press.

[2] E.V. Clark (1986): "The principle of contrast: a constraint on language acquisition" in B. MacWhinney(Ed.), "*Mechanisms of Language Acquisition*", Erlbaum.

[3] E.M.Markman (1987): "How children constrain the possible meanings of words" in U.Neisser(Ed.), "*Concepts and conceptual development*", Cambridge University Press.

[4] A. Takagi and Y. Ito (1987): "*Natural language processing*" (in Japanese), Maruzen.

[5] J. C. Hill(1983): "A model of language acquisition in the two-year-old", Cog. Brain Theory, vol.6, no.3, pp.287-317.

[6] M. Selfridge (1986):" A computer model of child language learning", Artificial Intelligence vol.29, pp.171-216.